Underepresentation of Black Individuals with Basal Cell Carcinoma (BCC) and Cutaneous Squamous Cell Carcinoma (cSCC) in U.S. Clinical Trials: A Comparison of the Demographics of Clinical Trial Participants to Those of an Insurance Claimant Cohort in the United States

Natalie Goulett

2024-09-15

# Summary

Non-melanoma Skin Cancers (NMSC), which are comprised of Basal Cell Carcinomas (BCC) and Cutaneous Squamous Cell Carcinomas (cSCC), are the most common cancers in the United States. To investigate the degree to which Black BCC and cSCC patients are represented in U.S. clinical trials, the proportions of Black participants in U.S. clinical trials for BCC and cSCC were compared to the proportions of Black individuals in a cohort who made insurance claims for BCC or cSCC treatment in the United States. Furthermore, this disparity was visualized and quantified by estimating how many Black individuals diagnosed in the United States with BCC or cSCC per year are unrepresented in U.S. clinical trials.

## Data Sources

A [cross-sectional study](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7450404/) by Lukowiak, et al. analyzed insurance claims data from the Optum Clinformatics Data Mart to compare the prevalence of BCC and cSCC between sexes, races, and geographic regions in the United States. The demographics of patients who made BCC and cSCC Medicare claims examined in this study are assumed to be reflective of the U.S. patient population and are assumed to represent the true BCC and cSCC patient demographics in the United States.

U.S. clinical trial demographic data were gathered from all publicly available clinical trials on BCC or cSCC conducted in the United States that were available on [ClinicalTrials.gov](https://clinicaltrials.gov/) as of June 28th, 2024.

In 2023, the [American Cancer Society](https://www.cancer.org/cancer/types/basal-and-squamous-cell-skin-cancer/about/key-statistics.html) estimated that the number of persons diagnosed with BCC or cSCC per year in the United States is 3,315,554. This number was used to estimate the numbers of Black BCC and cSCC patients diagnosed per year who are unrepresented in U.S. clinical trials.

## Methods

Two Pearson’s Chi-square tests with Yates’ continuity correction were conducted using RStudio to compare the proportions of Black participants that have participated in U.S. clinical trials vs. that of the commercially insured cohort for both BCC and cSCC cancers.

The Estimated number of persons diagnosed with skin cancer per year provided by the American Cancer Society, along with the demographics and race-specific ratio of BCC-to-cSCC diagnoses provided by Lukowiak, et al. and U.S. clinical trial demographic data, were used to estimate the numbers of Black BCC and cSCC patients diagnosed per year who are unrepresented in U.S. clinical trials.

# Chi-Square Test of U.S. BCC Clinical Trial Participants vs. Insurance Claimant Demographics

Out of 1,466 BCC clinical trial participants in the U.S., only 9 (0.6%) were Black. However, of the 618,516 individuals who made insurance claims for BCC treatment described in Lukowiak, et al., 23,315 (3.7%) were Black. A Pearson’s Chi-square test with Yates’ continuity correction was conducted to test the difference between the proportions of Black participants in these two groups.

## Packages Used

# Install and load required packages:  
# install.packages("tidyverse")  
# install.packages("extrafont") # optional  
  
library(tidyverse) # for data wrangling  
library(extrafont) # To use more fonts in graphs (optional)

## Data Transformation

# read BCC US clinical trial demographic data  
bccCtDemo\_df <- read\_csv(  
 "./Data/20240830\_bccUsctDemog.csv",  
 col\_names = TRUE,  
 na = c("#N/A", "#NA", "Not Reported", "Not reported"),  
 show\_col\_types = FALSE  
 ) %>%  
 rename(  
 "total\_participants" = "Number of participants analyzed",  
 "study\_ID" = "ClinicalTrials.gov ID",  
 "black\_participants" = "Black or African American",  
 "white\_participants" = "White"  
 ) %>%  
 select(  
 c(  
 "study\_ID",  
 "total\_participants",  
 "black\_participants",  
 "white\_participants"  
 )  
 ) %>%   
 filter(!is.na(black\_participants) | !is.na(white\_participants)  
)  
  
# calculate total participants  
bcc\_ct\_n <- sum(bccCtDemo\_df$total\_participants, na.rm = TRUE)  
# calculate number of Black participants  
bcc\_ct\_n\_black <- sum(bccCtDemo\_df$black\_participants, na.rm = TRUE)  
# calculate number of non-Black participants  
bcc\_ct\_n\_nonblack <- bcc\_ct\_n - bcc\_ct\_n\_black  
  
# Demographics from Lukowiak, et al.:  
# total number of claimants  
bcc\_claim\_n <- 618516  
# number of Black claimants  
bcc\_claim\_n\_black <- 23315  
# number of non-Black claimants  
bcc\_claim\_n\_nonblack <- bcc\_claim\_n - bcc\_claim\_n\_black  
  
# Create contingency table of expected and observed BCC patient demographics  
bcc\_demo\_contingency\_table <- matrix(  
 c(  
 bcc\_ct\_n\_nonblack,  
 bcc\_claim\_n\_nonblack,  
 bcc\_ct\_n\_black,  
 bcc\_claim\_n\_black  
 ),  
 nrow = 2  
)  
  
colnames(bcc\_demo\_contingency\_table) <- c(  
 "BCC Claimants",  
 "BCC Clinical Trial Participants"  
)  
rownames(bcc\_demo\_contingency\_table) <- c(  
 "Black Patients",  
 "Non-Black Patients"  
)

## Chi-Square Test (BCC)

# conduct chi square test to compare BCC clinical trial vs. insurance claimant  
# demographics  
bcc\_demo\_chi\_sq <- chisq.test(  
 bcc\_demo\_contingency\_table,  
 correct = TRUE  
)  
  
print(bcc\_demo\_chi\_sq)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: bcc\_demo\_contingency\_table  
## X-squared = 39.358, df = 1, p-value = 3.528e-10

### Interpretation

The proportion of insurance claimants with BCC who are Black (3.7%) is greater than the proportion of Black individuals enrolled in U.S. BCC clinical trials (0.6%) available on [ClinicalTrials.gov](https://clinicaltrials.gov/), (1, N = 619,982) = 39.358, *p* < 0.0001. This result strongly suggests the true proportion of Black BCC patients in the United States is greater than that of the respective U.S. clinical trial participants.

# Chi-Square Test of U.S. cSCC Clinical Trial Participants vs. Insurance Claimant Demographics

Out of 1,359 cSCC clinical trial participants in the U.S., only 19 (1.4%) were Black. However, of the 366,801 individuals who made insurance claims for cSCC treatment described in Lukowiak, et al., 16,070 (4.4%) were Black. A second Pearson’s Chi-square test with Yates’ continuity correction was conducted to test the difference between the proportions of Black participants in these two groups.

## Data Transformation

# read cSCC U.S. clinical trial demographic data  
csccCtDemo\_df <- read\_csv(  
 "./Data/20240829\_csccUsctDemog.csv",  
 col\_names = TRUE,  
 na = c("#N/A", "#NA", "Not Reported", "Not reported"),  
 show\_col\_types = FALSE  
 ) %>%  
 rename(  
 "total\_participants" = "Number of participants analyzed",  
 "study\_ID" = "ClinicalTrials.gov ID",  
 "black\_participants" = "Black or African American",  
 "white\_participants" = "White"  
 ) %>%  
 select(  
 c(  
 "study\_ID",  
 "total\_participants",  
 "black\_participants",  
 "white\_participants"  
 )  
 ) %>%   
 filter(!is.na(black\_participants) | !is.na(white\_participants)  
)  
  
# calculate total participants  
cscc\_ct\_n <- sum(csccCtDemo\_df$total\_participants, na.rm = TRUE)  
# calculate number of Black participants  
cscc\_ct\_n\_black <- sum(csccCtDemo\_df$black\_participants, na.rm = TRUE)  
# calculate number of non-Black participants  
cscc\_ct\_n\_nonblack <- cscc\_ct\_n - cscc\_ct\_n\_black  
  
# Demographics from Lukowiak, et al.:  
# total number of claimants  
cscc\_claim\_n <- 366801  
# number of Black claimants  
cscc\_claim\_n\_black <- 16070  
# number of non-Black claimants  
cscc\_claim\_n\_nonblack <- cscc\_claim\_n - cscc\_claim\_n\_black  
  
# Create contingency table of expected and observed cSCC patient demographics  
cscc\_demo\_contingency\_table <- matrix(  
 c(  
 cscc\_ct\_n\_nonblack,  
 cscc\_claim\_n\_nonblack,  
 cscc\_ct\_n\_black,  
 cscc\_claim\_n\_black  
 ),  
 nrow = 2  
)  
  
colnames(cscc\_demo\_contingency\_table) <- c(  
 "cSCC Claimants",  
 "cSCC Clinical Trial Participants"  
)  
rownames(cscc\_demo\_contingency\_table) <- c(  
 "Black Patients",  
 "Non-Black Patients"  
)

## Chi-Square Test (cSCC)

# conduct chi square test to compare cSCC clinical trial vs. insurance claimant  
# demographics  
cscc\_demo\_chi\_sq <- chisq.test(  
 cscc\_demo\_contingency\_table,  
 correct = TRUE  
)  
  
print(cscc\_demo\_chi\_sq)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: cscc\_demo\_contingency\_table  
## X-squared = 28.121, df = 1, p-value = 1.14e-07

### Interpretation

The proportion of insurance claimants with cSCC who are Black (4.4%) is greater than the proportion of Black individuals enrolled in U.S. cSCC clinical trials (1.4%) available on [ClinicalTrials.gov](https://clinicaltrials.gov/), (1, N = 368,160) = 28.121, *p* < 0.0001. This result strongly suggests the true proportion of Black cSCC patients in the United States is greater than that of the respective U.S. clinical trial participants.

# Black Representation in U.S. BCC and cSCC Clinical Trials

To investigate the degree of representation of Black BCC and cSCC patients in U.S. clinical trials, we will compare the estimated number of Black persons that will be diagnosed with BCC or cSCC in 2024 based on insurance claims data to the number of expected diagnoses if the BCC and cSCC insurance claimant population had the same demographic distribution as that of the participants in U.S. BCC and cSCC clinical trials.

# calculate proportion of BCC clinical trial participants who are Black  
bcc\_ct\_prop\_black <- bcc\_ct\_n\_black / bcc\_ct\_n  
# Estimated umber of new BCC and cSCC patients per year provided by the  
# American Cancer Society  
nmsc\_yr <- 3315554  
# proportion of NMSC claimants who are Black from Lukowiak, et al.:  
nmsc\_prop\_black <- 0.0413  
# proportion of BCC diagnoses among Black NMSC claimants from Lukowiak, et al.:  
prop\_black\_bcc <- 23315 / 39385  
# Estimate number of Black BCC patients diagnosed per year  
bcc\_n\_black\_yr <- nmsc\_yr \* nmsc\_prop\_black \* prop\_black\_bcc  
# calculate est Black BCC patients diagnosed per year that are represented by  
# clinical trial participants  
bcc\_n\_rep\_black\_yr <- nmsc\_yr \* bcc\_ct\_prop\_black \* prop\_black\_bcc  
# calculate number of unrepresented BCC patients diagnosed in 2024  
bcc\_n\_unrep\_black\_yr <- bcc\_n\_black\_yr - bcc\_n\_rep\_black\_yr  
  
# calculate proportion of cSCC clinical trial participants who are Black  
cscc\_ct\_prop\_black <- cscc\_ct\_n\_black / cscc\_ct\_n  
# proportion of cSCC diagnoses among Black NMSC claimants from Lukowiak, et al.:  
prop\_black\_cscc <- 16070 / 39385  
# Estimate number of Black cSCC patients diagnosed per year  
cscc\_n\_black\_yr <- nmsc\_yr \* nmsc\_prop\_black \* prop\_black\_cscc  
# calculate est Black cSCC patients diagnosed per year that are represented by  
# clinical trial participants  
cscc\_n\_rep\_black\_yr <- nmsc\_yr \* cscc\_ct\_prop\_black \* prop\_black\_cscc  
# calculate number of unrepresented cSCC patients diagnosed in 2024  
cscc\_n\_unrep\_black\_yr <- cscc\_n\_black\_yr - cscc\_n\_rep\_black\_yr  
  
# proportion of Black patients unrepresented  
bcc\_n\_unrep\_black\_yr / bcc\_n\_black\_yr

## [1] 0.8513522

cscc\_n\_unrep\_black\_yr / cscc\_n\_black\_yr

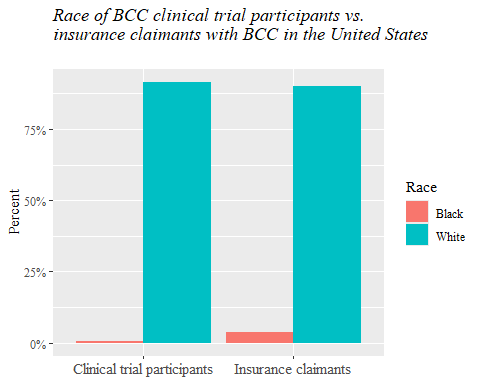
## [1] 0.6614802

It was estimated that among Black individuals in the United States, 81,061 will be diagnosed with BCC and 55,872 will be diagnosed with cSCC in 2024. Accounting for the percent of participants who are Black in BCC (0.6%) and cSCC (1.4%) clinical trials, it was estimated that only 12,049 (14.9%) of these Black individuals diagnosed with BCC and 18,914 (33.9%) diagnosed with cSCC per year will be represented by clinical trial participants. As a result, an estimated 69,011 (85.1%) of these Black individuals diagnosed with BCC and an estimated 36,958 (66.1%) diagnosed with cSCC will not be represented in U.S. clinical trials.

## Visualization of Demographics in U.S. Clinical Trials vs. U.S. Population

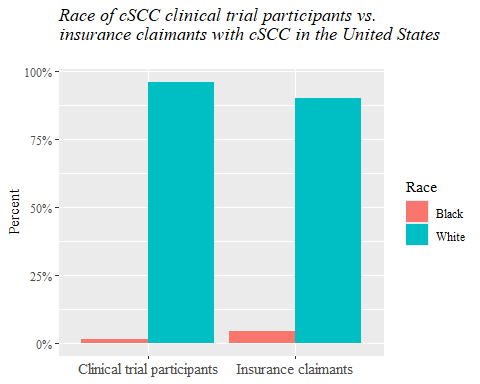
The percentages of White and Black individuals in BCC clinical trials vs. that of insurance claimants with BCC are visually compared below:

# loadfonts(device = "win")  
  
# proportion of US BCC patients who are black  
bcc\_claim\_prop\_black <- bcc\_claim\_n\_black / bcc\_claim\_n  
# calculate proportions of white and black BCC clinical trial participants  
bcc\_ct\_n\_white <- sum(bccCtDemo\_df$white\_participants, na.rm = TRUE)  
bcc\_ct\_prop\_white <- bcc\_ct\_n\_white / bcc\_ct\_n  
# calculate proportions of white and black BCC claimants  
bcc\_claim\_n\_white <- 556716  
bcc\_claim\_prop\_white <- bcc\_claim\_n\_white / bcc\_claim\_n  
  
# barplot of % white and black participants in BCC clinical trials vs claims  
bar\_pop <- c(  
 rep("Clinical trial participants", 2),  
 rep( "Insurance claimants", 2)  
)  
bar\_race <- rep(c("Black", "White"), 2)  
bcc\_prop <- c(  
 bcc\_ct\_prop\_black,  
 bcc\_ct\_prop\_white,  
 bcc\_claim\_prop\_black,  
 bcc\_claim\_prop\_white  
)  
bcc\_demoBar <- data.frame(  
 bar\_pop,  
 bar\_race,  
 bcc\_prop  
)  
  
bccplot <- ggplot(  
 bcc\_demoBar,  
 aes(  
 fill = bar\_race,  
 y = bcc\_prop,  
 x = bar\_pop  
 )  
 ) +  
 geom\_bar(position = "dodge", stat = "identity") +  
 labs(  
 title = str\_wrap(  
 "Race of BCC clinical trial participants vs. insurance claimants with BCC in the United States",  
 width = 50  
 ),  
 fill = "Race",  
 y = "Percent",  
 x = NULL  
 ) +  
 scale\_y\_continuous(labels = scales::percent) +  
 theme(  
 text = element\_text(family = "Times New Roman"),  
 axis.text.x = element\_text(size = 11),  
 axis.title.y = element\_text(size = 11),  
 title = element\_text(size = 11),  
 plot.title = element\_text(  
 hjust = 0,  
 face = "italic",  
 margin = margin(b = 20)   
 )  
 )  
  
print(bccplot)



Likewise, the percentages of White and Black individuals in cSCC clinical trials vs. that of insurance claimants with cSCC are visually compared below:

# proportion of US cSCC patients who are black  
cscc\_claim\_prop\_black <- cscc\_claim\_n\_black / cscc\_claim\_n  
# calculate proportions of white and black BCC clinical trial participants  
cscc\_ct\_n\_white <- sum(csccCtDemo\_df$white\_participants, na.rm = TRUE)  
cscc\_ct\_prop\_white <- cscc\_ct\_n\_white / cscc\_ct\_n  
# calculate proportions of white and black BCC claimants  
cscc\_claim\_n\_white <- 330285  
cscc\_claim\_prop\_white <- cscc\_claim\_n\_white / cscc\_claim\_n  
  
# barplot of % white and black participants in BCC clinical trials vs claims  
cscc\_prop <- c(  
 cscc\_ct\_prop\_black,  
 cscc\_ct\_prop\_white,  
 cscc\_claim\_prop\_black,  
 cscc\_claim\_prop\_white  
)  
cscc\_demoBar <- data.frame(bar\_pop, bar\_race, cscc\_prop)  
  
csccplot <- ggplot(  
 cscc\_demoBar,  
 aes(  
 fill = bar\_race,  
 y = cscc\_prop,  
 x = bar\_pop  
 )  
 ) +  
 geom\_bar(position = "dodge", stat = "identity") +  
 labs(  
 title = str\_wrap(  
 "Race of cSCC clinical trial participants vs. insurance claimants with cSCC in the United States",  
 width = 50  
 ),  
 fill = "Race",  
 y = "Percent",  
 x = NULL  
 ) +  
 scale\_y\_continuous(labels = scales::percent) +  
 theme(  
 text = element\_text(family = "Times New Roman"),  
 axis.text.x = element\_text(size = 11),  
 axis.title.y = element\_text(size = 11),  
 title = element\_text(size = 11),  
 plot.title = element\_text(  
 hjust = 0,  
 face = "italic",  
 margin = margin(b = 20)  
 )  
 )  
  
print(csccplot)



As seen in both graphs, lower proportions of Black individuals and higher proportions of White individuals are seen in both BCC and cSCC clinical trials when compared to the commercially insured cohort.